REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
MAY 2013	JOURNAL ARTICLE (Post Print)	DEC 2010 – NOV 2012

4. TITLE AND SUBTITLE

THE 3D STACKING BIPOLAR RRAM FOR HIGH DENSITY

5a. CONTRACT NUMBER
FA8750-11-2-0046

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
62788F

6. AUTHOR(S)

Yi-Chung Chen, Hai Li, Wei Zhang, Robinson Pino

5d. PROJECT NUMBER
T2NC

5e. TASK NUMBER
PO

5f. WORK UNIT NUMBER
LY

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Polytechnic Institute of NYU
2 MetroTech Center
Brooklyn, NY 11201

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/Information Directorate
Rome Research Site/RITB
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TP-2013-018

12. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA Case Number: 88ABW-2011-3396

13. SUPPLEMENTARY NOTES

14. ABSTRACT
For its simple structure, high density and good scalability, the resistive random access memory (RRAM) has emerged as one of the promising candidates for large data storage in computing systems. Moreover, building up RRAM in a three dimensional (3D) stacking structure further boosts its advantage in array density. Conventionally, multiple bipolar RRAM layers are piled up vertically separated with isolation material to prevent signal interference between the adjacent memory layers. The process of the isolation material increases the fabrication cost and brings in the potential reliability issue. To alleviate the situation, we introduce two novel 3D stacking structures built upon bipolar RRAM crossbars that eliminate the isolation layers. The bi-group operation scheme dedicated for the proposed designs to enable multi-layer accesses while avoiding the overwriting induced by the cross-layer disturbance, is also presented. Our simulation results show that the proposed designs can increase the capacity of a memory island to 8K-bits (i.e., 8 layers of 32 x 32 crossbar array) while maintaining the sense margin in the worst-case configuration greater than 20% of the maximal sensing voltage.

15. SUBJECT TERMS
Resistive memory, RRAM, 3D stacking, crossbar, bipolar operation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			NATHAN MCDONALD
U	U	U	UU	10	19b. TELEPHONE NUMBER (*Include area code*) N/A

The 3D Stacking Bipolar RRAM for High Density

Yi-Chung Chen, Hai (Helen) Li, *Member, IEEE,* Wei Zhang, *Member, IEEE,* Robinson E. Pino *Senior Member, IEEE*

***Abstract*—For its simple structure, high density and good scalability, the resistive random access memory (RRAM) has emerged as one of the promising candidates for large data storage in computing systems. Moreover, building up RRAM in a three-dimensional (3D) stacking structure further boosts its advantage in array density. Conventionally, multiple bipolar RRAM layers are piled up vertically separated with isolation material to prevent signal interference between the adjacent memory layers. The process of the isolation material increases the fabrication cost and brings in the potential reliability issue. To alleviate the situation, we introduce two novel 3D stacking structures built upon bipolar RRAM crossbars that eliminate the isolation layers. The bi-group operation scheme dedicated for the proposed designs to enable multi-layer accesses while avoiding the overwriting induced by the cross-layer disturbance, is also presented. Our simulation results show that the proposed designs can increase the capacity of a memory island to 8K-bits (*i.e.*, 8 layers of 32×32 crossbar array) while maintaining the sense margin in the worst-case configuration greater than 20% of the maximal sensing voltage.**

***Index Terms*—Resistive memory, RRAM, 3D stacking, crossbar, bipolar operation.**

I. Introduction

As the conventional memory technologies, *e.g.*, SRAM, DRAM and Flash, are approaching their physical limitations, the rapidly increasing technology difficulties and fabrication costs encourage researchers to look for other replacements [1]. Many emerging non-volatile memories, such as the *resistive RAM* (RRAM) [2], the *phase-change RAM* (PC-RAM) [3], and the *spin-transfer torque RAM* (STT-RAM) [4], have been investigated and prototyped in recently years. Among them, RRAM has become a promising candidate to substitute the traditional data storage technologies, *e.g.*, hard disk drive and flash memory, for its high density, low power consumption, and good scalability [5].

In general, RRAM can be used to denote all the random access memories that rely on the resistance differences to store data. Various materials based on the different physical mechanisms have been extensively studied [6]. Recently, RRAM technologies has also been extended to build memristors [7][8] – the 4^{th} fundamental passive circuit element [9]. From the programming point of view, these technologies can be simply cataloged into either *unipolar* or *bipolar* switching. In this work, we target density and bandwidth improvement of the bipolar RRAM memories.

The *three-dimensional* (3D) stacking that builds up multiple memory layers vertically is an efficient way for density improvement. Conventionally, the isolation layer is inserted between two adjacent memory layers to avoid the malfunctions caused by the signal interference when simultaneously accessing multiple memory layers [10]. Manufacturing the isolation layers could introduces potential reliability issues, such as the melting (or even destruction) of metal interconnects during the annealing step. Therefore, the low thermal budget process with high process complexity and fabrication cost, *e.g.*, *undoped Methylsilsesquioxane* (MSQ) *Spin-on-Glass* (SOG) technology, is needed [11][12].

To relax the requirements of fabrication process and enhance the density for bipolar RRAM design, we propose two 3D stacking structures by utilizing the same design concept. The two structures are named as *3D Interleaved Complementary Memory Layer* (3D-ICML) and *3D High-density Interleaved Memory* (3D-HIM). Our designs consist of two types of material stacks: one has the original deposition sequence and the other is deposited in the reversed order. A memory island is formed by applying these two material stacks to odd and even layers alternately and by sharing the electrodes and interconnection metal wires between two adjacent layers. As we shall demonstrate in Section IV, the proposed interleaved 3D stacking structures can be built upon crossbar arrays without isolation layers. Furthermore, with the aid of our proposed bi-group operation scheme, these designs can obtain a high throughput by simultaneously accessing memory cells on the different layers without signal interference or unwanted overwriting.

Our simulations show that the stacking structures can function properly for a memory island consisting of 8 layers of 32×32 crossbar arrays. The minimal sensing margin is more than 20% of the maximal sensing voltage, which can be easily read out by the peripheral circuit. The related design implications on the impacts of sneak paths on 3D crossbar arrays and the design constraints in read and write operations have also been discussed and explored.

The rest of the paper is organized as follows: Section II gives a preliminary introduction on the RRAM device's fundamental and the crossbar array. Section III summarizes the previous work on 3D bipolar RRAM designs and explains the process difficulties. In Section IV, we describe the design concept of the proposed stacking structures and the bi-group operation scheme in read and write operations. The simulation results and discussion on the design implications are presented in Section V. At the end, we conclude the paper in Section VI.

Yi-Chung Chen and Hai (Helen) Li are with the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, Brooklyn, New York, USA. (email: ychen33@student.poly.edu, hli@poly.edu)

Wei Zhang is with the School of Computer Engineering, Nanyang Technological University, Singapore. (email: zhangwei@ntu.edu.sg)

Robinson E. Pino is with the Network Science Division, ICF International, Fairfax, Virginia, USA. (email: robinson.pino@icfi.com)

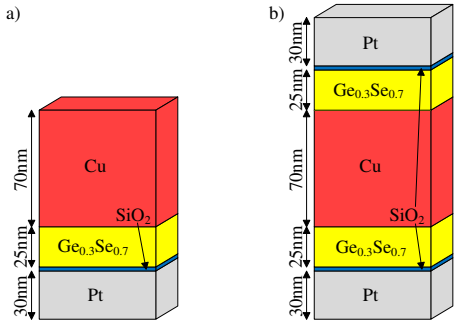


Fig. 1. (a) Structure of Cu-$Ge_{0.3}Se_{0.7}$-SiO_2-Pt [13]. (b) The complementary cell structure [14].

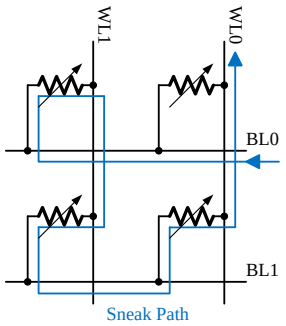


Fig. 2. RRAM crossbar array and the sneak path.

II. Preliminary

A. Resistive Random Access Memory

The *resistive random access memory* (RRAM) can be realized by many different materials based on the different storage mechanisms. All of these materials fall into only two operation types – *unipolar* switching and *bipolar* switching. Within this context, unipolar operation executes the programming/erasing by using short/long pulses, or by using high/low voltage with the same voltage polarity. In contrast, bipolar switching is implemented by short pulses with opposite voltage polarity for programming and erasing [15]. In this work, we target mainly on 3D structures with bipolar RRAM devices for their fast switching speed and the less power consumption in *RESET* (that is, erase) operation [16]. For demonstration purpose, we use the material Cu-$Ge_{0.3}Se_{0.7}$-SiO_2-Pt [13] as example. However, the proposed design concept can be easily extended to the other bipolar RRAM devices.

Fig. 1(a) illustrates the structure of Cu-$Ge_{0.3}Se_{0.7}$-SiO_2-Pt [13]. It is a programmable metallization cell device formed in a sandwich structure with heterogeneous solid metal electrodes at two poles. One pole is relatively inert Pt (called as the *bottom electrode*, or BE), the other is electro-chemically active Cu (called as the *top electrode*, or TE). A thin electrolyte film composed of ternary glass $Ge_{0.3}Se_{0.7}$ with added dissolved active metal Cu is placed between the two electrodes. The SiO_2 is used as a buffer layer to improve the endurance in the electrolyte [17]. The $Ge_{0.3}Se_{0.7}$ and SiO_2 are the places where the resistance changing happens.

For convenience, we define $\mathrm{R_{on}}$ and $\mathrm{R_{off}}$ as the resistance value at the *low resistance state* (LRS) and at the *high resistance state* (HRS), respectively. The $\mathrm{R_{off}/R_{on}}$ is an important device parameter representing the difference between HRS and LRS. In general, a high $\mathrm{R_{off}/R_{on}}$ is more preferable.

When a negative bias is applied to the BE during a *SET* operation (that is, the device changes to the LRS), the dissolving Cu reacts with Se in electrolyte compound to generate cation conductors which forms a "filament" between two electrodes for electron transportation. As a result, the resistance between two electrodes is dramatically reduced. To *RESET* a cell (to change the device to the HRS), a positive bias can be applied on the BE and remove the random dissolving Cu from Cu-Ge-Se compound filament. The resistance becomes relatively high once the filament disappears in the electrolyte [17].

B. Crossbar Array

Crossbar array is widely used in RRAM design for its simple structure and high density. Crossbar was firstly initiated and demonstrated in a telecommunication switching system, which contained two sets of wires and switches at cross points. Signal routing is controlled by properly selecting switches. In the nanometer-scale high-density memory design, the similar structure is maintained – a storage element is placed at each cross point of two sets of metal wires [18]. Theoretically, using crossbar array structure can achieve the smallest memory cell area $4\mathrm{F}^2$, where F is the minimum feature size [15].

However, the crossbar array also results in sneak path in which three or more cells are connected in series as shown in Fig. 2. To guarantee the proper functionalities in both write and read operations, the voltage/current across the selected memory cell must be much higher than the overall amount of current absorbed by the unselected cells [15]. On the other hand, the voltage across an unselected cell must be smaller than the threshold of the SET/RESET operation to avoid the unwanted resistance change. To control the impacts of sneak paths within an acceptable range, the size and hence the capacity of a crossbar array is limited.

C. The Complementary RRAM

Recently, Linn *et al.* proposed a complementary RRAM cell structure, which is made of two anti-serial RRAM devices as illustrated in Fig. 1(b) [14]. Under all the possible operation conditions, at least one of the two RRAM devices in this complementary cell exhibits the HRS state, which can dramatically reduce the impact of sneak paths. However, any single data recording has to be associated with a multi-step write procedure which requires a careful and complex operation configuration. This design also brings in severe issues in terms of the high power consumption and the degraded device reliability. Moreover, considering that each memory cell includes two complementary RRAM devices, the memory capacity is only half of a conventional 3D RRAM design.

III. Related Work

Simply stacking multiple memory layers vertically is a common way to construct 3D design with bipolar RRAM devices [19]. Each memory layer has its own set of storage elements and interconnects. An isolation layer is inserted between two neighboring layers to prevent the signal interference.

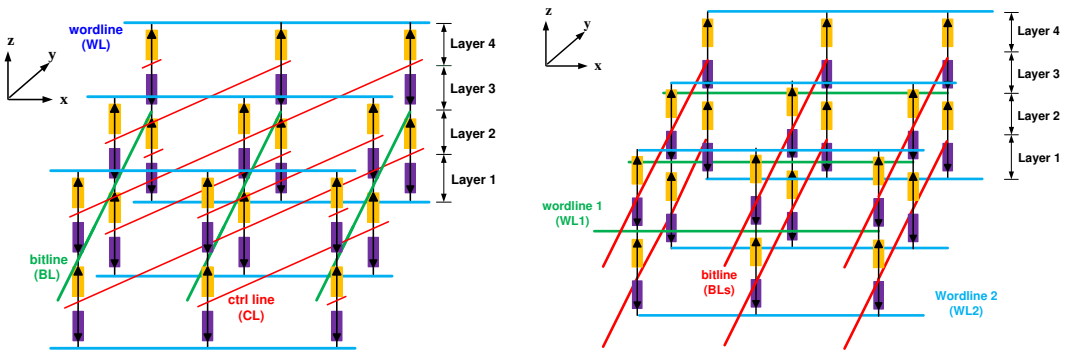


Fig. 3. The proposed 3D stacking structures for bipolar RRAM: (a) 3D-ICML; and (b) 3D-HIM.

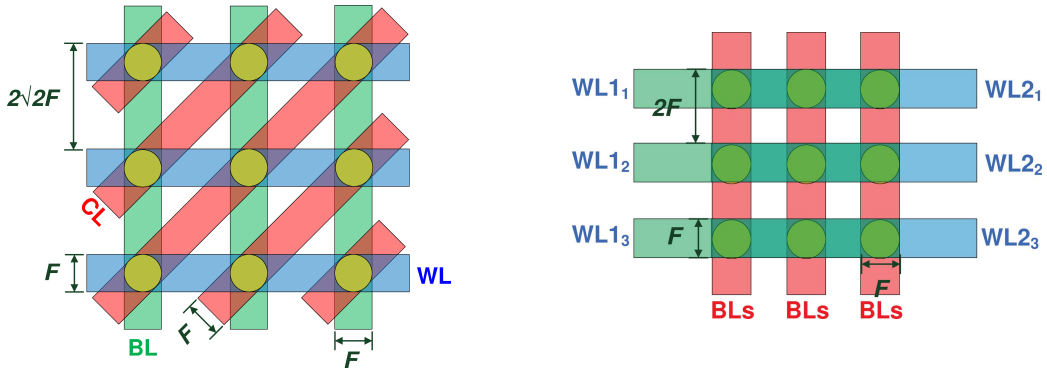


Fig. 4. The layouts of the proposed 3D stacking structures: (a) 3D-ICML; and (b) 3D-HIM.

Recently, an improved design was proposed by Kugeler *et al.*, which the *word lines* (WLs) between two memory layers can be shared [20]. The two memory layers sharing the same WLs can be accessed and programmed simultaneously. However, *bit lines* (BLs) cannot be shared, and hence, the manufacturing of isolation layers are still needed.

SOG with MSQ *etc.* materials can be used to form isolation layers. However, there are some critical difficulties from a process development point of view, including device degradation due to thermal processing [21], misalignment of vias due to SOG [22], poor adhesion of SOG material [11], and heat accumulation because of the low conductivity of the isolation material [21][23]. Consequently, a 3D memory design excluding isolation process is preferred for lower fabrication cost and process complexity. Previously, Jonson *et al.* presented a bipolar multi-layered conductive metal oxide memory without isolation layer, but it can be applied only to one-time programming ROM applications [24].

IV. The Proposed 3D Stacking Structures

In this work, we propose two 3D stacking structures for bipolar RRAMs, namely, *3D Interleaved Complementary Memory Layer* (3D-ICML) and *3D High density Interleaved Memory* (3D-HIM). This section describes the proposed designs and explains the bi-group operation scheme for throughput improvement.

A. 3D-ICML and 3D-HIM

Fig. 3(a) and Fig. 3(b) illustrate 3D-ICML and 3D-HIM, respectively. For simplicity, we demonstrate only four memory layers, each of which consists of a RRAM crossbar array.

The basic design concept is to employ the complementary material stacking structures, *i.e.*, the regular memory stack and the one with a reversed deposition order, to the memory cells in the adjacent layers. For instance, all the memory cells of Layer 1 use the regular deposition process (illustrated with *purple pillars* in the figures), and those of Layer 2 are made by reversing the deposition sequence (*yellow pillars*). The two types of memory stacks are applied to the odd and the even layers alternatively. This process has been successfully demonstrated by Linn *et al.*. And their experiments showed that the memory cells made with regular and reversed depositions can provide the same device properties [14].

In both 3D-ICML and 3D-HIM, memory devices and metal wires form a memory island without isolation layers. Any two adjacent memory cells at the same (x, y) location are connected back to back, and hence, share the metal wire in between. However, the signal sharing schemes among memory layers in the two designs are different, as shown in Fig. 3(a) and Fig. 3(b). In 3D-ICML, the Pt electrodes are connected to either WLs along the $x-axis$ or BLs along $y-axis$. A new set of wires connected to the Cu electrodes are introduced and routed along the diagonal direction to $x-axis$ and $y-axis$. On the contrary, 3D-HIM does not have diagonal routings: the Pt electrodes are connected to WL1s or WL2s along the $x-axis$, and BLs along the $y-axis$ contact to the Cu electrodes. The difference of 3D-ICML and 3D-HIM also reflect on their corresponding layouts in Fig. 4(a) and Fig. 4(b), respectively.

TABLE I
Cu-$Ge_{0.3}Se_{0.7}$-SiO_2-Pt & SOG PROCESS PARAMETERS [11][17]

Material	Cu	$Ge_{0.3}Se_{0.7}$	SiO_2	Pt	SOG
Thickness (nm)	70	25	3	30	50

TABLE II
VARIOUS RRAM PROCESS PARAMETERS

Device	Conventional 3D RRAM	3D-ICML	Density Improve
Cu-$Ge_{0.3}Se_{0.7}$-SiO_2-Pt [14]	$178nm \cdot 4F^2$	$78nm \cdot 8F^2$	14.1%
Pt-TiO_2-Pt [25]	$130nm \cdot 4F^2$	$55nm \cdot 8F^2$	18.1%
Au-ZrO_2-Ag [26]	$190nm \cdot 4F^2$	$90nm \cdot 8F^2$	5.6%
Ni/GeO/STO/TaN [27]	$145nm \cdot 4F^2$	$57.5nm \cdot 8F^2$	26.1%

TABLE III
VARIOUS RRAM PROCESS PARAMETERS

Device	Conventional 3D RRAM	3D-HIM	Density Improve
Cu-$Ge_{0.3}Se_{0.7}$-SiO_2-Pt [14]	$178nm \cdot 4F^2$	$78nm \cdot 4F^2$	128%
Pt-TiO_2-Pt [25]	$130nm \cdot 4F^2$	$55nm \cdot 4F^2$	136%
Au-ZrO_2-Ag [26]	$190nm \cdot 4F^2$	$90nm \cdot 4F^2$	111%
Ni/GeO/STO/TaN [27]	$145nm \cdot 4F^2$	$57.5nm \cdot 4F^2$	152%

B. Memory Density Comparison

As shown in Fig. 4(b), each array layer in 3D-HIM is exactly the same as a conventional crossbar. Therefore, the smallest cell area that 3D-HIM can obtain is

$$A_{\mathrm{3D-HIM}} = A_{\mathrm{conv}} = 4F^2.$$

The cell area of 3D-ICML is constrained by the space between CLs so that the space between RRAM devices cannot keep at the minimum. Accordingly,

$$A_{\mathrm{3D-ICML}} = [(2\sqrt{2})F]^2 = 8F^2.$$

Note that for a 3D memory, its density is determined by not only the single memory cell area, but also the allowable number of memory layers. By sharing BEs and TEs among neighboring layers, our design can reduce the overall number of conduction layers and remove isolation layers. For a given height of a 3D structure, which usually is a major limitation in fabrication process, more memory layers can be stacked up vertically. Thus, the memory capacity increases.

Table I lists the geometric parameters of Cu-$Ge_{0.3}Se_{0.7}$-SiO_2-Pt and SOG. In the conventional 3D RRAM structure, a layer of memory array is constructed using one memory device and a SOG layer. The total thickness is approximate $178nm$. In contrast, a memory layer of the proposed structures has a thickness of only $78nm$. We can define the memory density as

$$D = 1/(A \cdot T),$$

where, A and T represent a single memory cell area and memory layer thickness, respectively. Accordingly, we have

$$D_{\mathrm{3D-conv}} = 1/(4F^2 \cdot 178nm),$$

$$D_{\mathrm{3D-ICML}} = 1/(8F^2 \cdot 78nm), \text{ and}$$

$$D_{\mathrm{3D-HIM}} = 1/(4F^2 \cdot 78nm).$$

Compared to the conventional 3D RRAM, 3D-ICML can increase the memory density 14.1% even it has a bigger memory cell. The density enhancement obtained by 3D-HIM is even more than 100%. More examples for the different RRAM materials can be found in Table II and III.

C. Bi-Group Operation Scheme

To enable the simultaneous access to the RRAM cells in multiple layers and therefore enhance the data throughput in read and write operations, we propose the *Bi-Group Operation Scheme*, which can be applied to both 3D-ICML and 3D-HIM. In this subsection, we describe the design principle by using only 3D-HIM as an example due to the space limitation. For ease of discussion, let's define some terms used in 3D-HIM.

- **WL1s** and **WL2s**: we number the WL layer at the bottom of the 3D-HIM structure as '0' and continue counting the other WL layers from bottom to top. WL1s (WL2s) represent those WL layers with odd (even) numbers.
- **WL1$_i$GC** and **WL2$_j$GC**: we name the group of memory cells connected to a given $WL1_i$ or $WL2_j$ as $WL1_i$GC (that is, $WL1_i$ group cells) or $WL2_j$GC, respectively.

Totally, three sets of control signals, *i.e.*, BL, WL1 and WL2, are utilized. Each of them is responsible to the related operations on the memory layers above and below it.

1) Overview of the Bi-Group Operation Scheme: In 3D-HIM, there are two sets of group cells – $WL1_i$GC and $WL2_j$GC. Only one of them can be accessed at once during read or write operations. This scheme has several advantages: (1) It increases throughput by simultaneously accessing multiple memory cells within either $WL1_i$GC or $WL2_j$GC. (2) The unselected groups can be biased to ground and taken as the signal isolators. Thus, the unexpected overwriting caused by the write operations on different memory layers can be eliminated. (3) The BLs are shared by the RRAM layers above the BLs, and below BLs. The peripheral circuitry connected to the BLs are also shared by two RRAM layers to reduce area cost. Furthermore, WL1s and WL2s can be driven from the opposite sides of the memory island as shown in Fig. 3(b) to distribute the layouts of peripheral circuitry.

2) Read Operation: When reading out the stored data in a RRAM cell, we provide a sense voltage (V_{sense}) to the corresponding WL, and measure the current through the cell. To prevent the unexpected overwriting, V_{sense} should be much smaller than the threshold voltage to switch the RRAM device. A sense amplifier is connected to the BL and shared by two group's cells $WL1_i$GC and $WL2_i$GC. Based on the bi-group operation scheme, only a set of group's cells can be sensed out at one time.

Fig. 5(a) shows an example of reading out the cells in $WL1_1$GC. Accordingly, $WL1_1$ is raised to V_{sense} and all the other $WL1_i$ are tied to 0 V. To prevent the disturbance from/to WL2 groups, all the WL2s are forced to 0 V. Similarly, the read operation of $WL2_j$GC on the $x-y$ plane can be accessed simultaneously (which is omitted here due to space restriction).

An active load (R_{sense}) is used at the end of BL to transfer current through the memory device to the input voltage of a sense amplifier $V_{\mathrm{R-sense}}$. To simplify the evaluation of the read operation in this work, $R_{\mathrm{sense}} = 100\Omega$) is used as the input resistance of sense amplifier. The *sensing margin* (SM) is defined as the ratio of $V_{\mathrm{R-sense}}$ with respect to V_{sense}.

We use $N \times N \times H$ to represent a memory island with H layers and each layer has a $N \times N$ crossbar array. The corresponding *memory capacity* is $MC = N^2 \cdot H$ and the *read*

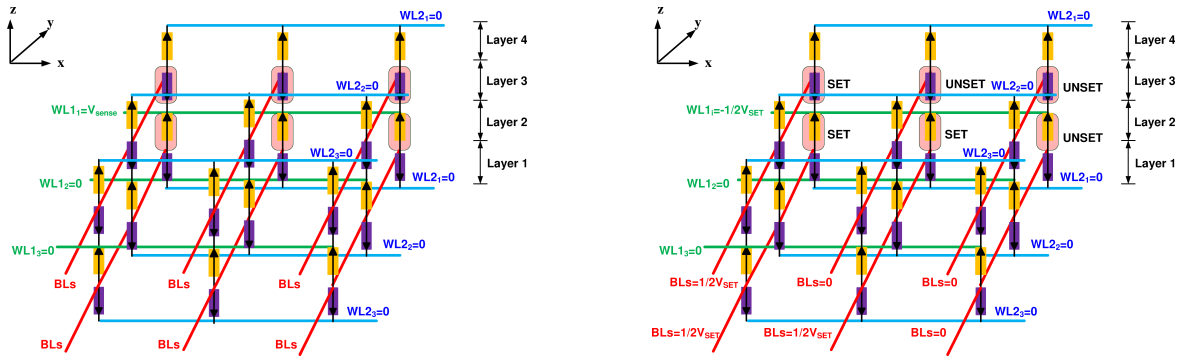


Fig. 5. The selected WL_1GC of 3D-HIM (a) in read operation; and (b) in SET operation.

bandwidth is $\mathrm{BW_{Read}} = \mathrm{N} \cdot \mathrm{H}/2$.

3) Write Operation: The two types of write procedures (SET and RESET) for bipolar RRAM require the opposite driving polarities and hence have to be separated. Combined with the bi-group operation scheme, 3D-HIM can simultaneously program the memory cells that locate in the same group and have the same incoming data.

The driving conditions need to be carefully controlled to avoid unexpected overwriting caused by sneak paths and to minimize the total write current. Table IV summarizes the ideal biasing voltages when performing SET and RESET operations. All the other WL1s, WL2s and BLs that are not related to the current writing operation are forced to 0 V.

Fig. 5(b) illustrates an example of $\mathrm{WL1}_1$GC during a SET operation. For illustration purpose, we assume half of the cells in $\mathrm{WL1}_1$GC are programmed at the same time. The $\mathrm{WL1}_1$ are biased to $-0.5\mathrm{V_{SET}}$, the BLs connected to the cells to be programmed are forced to $0.5\mathrm{V_{SET}}$, and all the unrelated control signals are set to $0V$. As shown in Fig. 5(b), the voltage drop across an unselected cell within $\mathrm{WL1}_1$GC is only $0.5\mathrm{V_{SET}}$, which is too small to change its resistance state. The RESET procedure works similarly though the biasing condition is different.

When all the cells in the given group are programmed to the same value, 3D-HIM obtains the maximal *write bandwidth* $\mathrm{BW_{max}} = \mathrm{N} \cdot \mathrm{H}/2$. The average write bandwidth is $\mathrm{BW_{avg}} = \mathrm{N} \cdot \mathrm{H}/4$ since a memory cell is either set or reset.

V. Simulation Result and Discussion

This section discusses the major design issues of the conventional 3D stacking crossbar arrays induced by sneak paths and demonstrate how the proposed structure effectively alleviate the situation and improve memory density.

TABLE IV
Driving Conditions of Writing Operations

Data	Cell Group	Driving Conditions
LRS	$\mathrm{WL1}_i$GC	WL1: $-0.5 \cdot \mathrm{V_{SET}}$, BL: $0.5 \cdot \mathrm{V_{SET}}$
LRS	$\mathrm{WL2}_j$GC	WL2: $-0.5 \cdot \mathrm{V_{SET}}$, BL: $0.5 \cdot \mathrm{V_{SET}}$
HRS	$\mathrm{WL1}_i$GC	WL1: $-0.5 \cdot \mathrm{V_{RESET}}$, BL: $0.5 \cdot \mathrm{V_{RESET}}$
HRS	$\mathrm{WL2}_j$GC	WL2: $-0.5 \cdot \mathrm{V_{RESET}}$, BL: $0.5 \cdot \mathrm{V_{RESET}}$

TABLE V
RRAM Parameters

Parameters	Value	Parameters	Value
$\mathrm{V_{SET}}$	1.5 V	$\mathrm{R_{IR}}$	2.5 Ω
$\mathrm{V_{RESET}}$	1 V	$\mathrm{V_{sense}}$	0.1 V
$\mathrm{R_{off}}$(HRS)	1 MΩ	$\mathrm{R_{sense}}$	100 Ω
$\mathrm{R_{on}}$(LRS)	5 kΩ		

A. Simulation Setup

Table V summarizes the RRAM parameters used in this work [28]. In a crossbar array, the *interconnect resistance* (IR) along the driving path results in voltage drop and decreases the driving voltage delivered to the target memory cells. Therefore, we integrated the IR in our simulation model and set the IR per memory cell $\mathrm{R_{IR}} = 2.5\Omega$, estimated based on the DRAM interconnect data at $22nm$ technology node [1]. We assume up to eight memory layers in the stacking structure after considering the process limitation. All the simulations were conducted by using Spectre on Cadence CAD platform.

B. Impact of Data Pattern and Cell Location

1) Impact of Data Pattern: The effectiveness of read and write operations in a stacking memory island is related to the data distribution within the structure, or, *data pattern*. We divide all the memory cells into three categories determined by their location: the target cell, the cells along the driving path (*i.e.*, $\mathrm{WL1}_1$ or $\mathrm{WL2}_1$), and all the other cells. An example of the $\mathrm{WL1}_1$GC of 3D-HIM in the read operation is shown in Fig. 6. Here, the target cell is highlighted in color *red*, and the cells along the driving path WL1 are marked with color *blue*.

Four typical data patterns – "LL", "LH", "HH" and "HL" – are usually used in RRAM crossbar array analysis. Here, the first and the second letters represent the resistance status of the target cell and those on the driving path, respectively ('L' represents LRS, and 'H' is HRS). By default, we assume all the other cells are at LRS, which corresponds to the worse-case configuration in read and write operations. The definition of the data pattern can also be applied to 3D-ICML, which will be used in the following discussion.

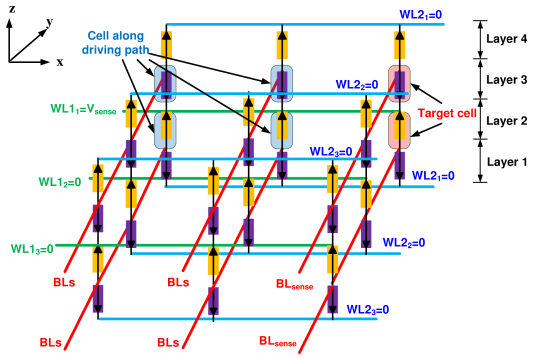


Fig. 6. 3D-HIM memory cell classification for data pattern analysis.

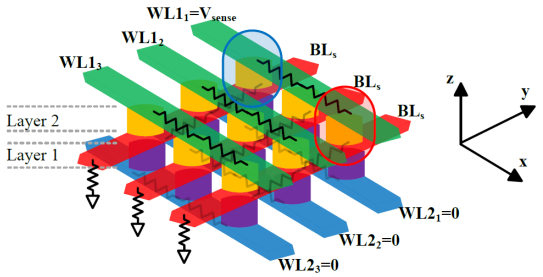


Fig. 7. The worst-case and the best-case scenarios in terms of cell location.

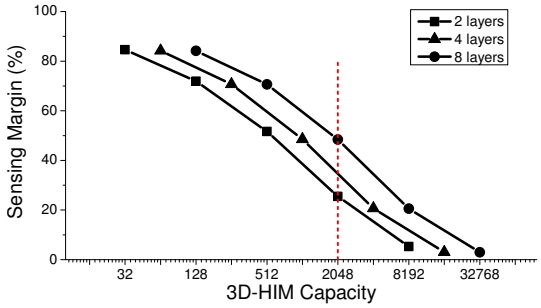


Fig. 9. SM of 3D-HIM versus capacity of one memory island.

2) Impact of Cell Location: The physical location of the target cell within an array also affects its operation scenario. Fig. 7 illustrates a two-layer 3D-HIM in a read operation. In this example, the driving current flows from the leftmost side of $WL1_1$ to the rightmost of the array along $x-axis$. Because of the existence of interconnect resistance, the voltage applied on the memory cells along $WL1_1$ are not the same. The cell in the right corner (highlighted in color *red*) suffers most from the voltage drop on $WL1_1$ and hence has the smallest sense margin. In the contrast, the best situation happens to the cell in the left corner (highlighted in color *blue*), which is affected least by the interconnect resistance.

Fig. 8(a) shows the SM difference between the worst-case and the best-case cell locations in a four-layer 3D-HIM. Comparing the four typical data patterns, the crossbar array size demonstrates the biggest impact on the 'LL' pattern: the SM difference incurred by only the location could be $\sim 10\%$ of the maximal sensing voltage. This is because a target cell at LRS results in a large driving current and hence a high voltage drop on the interconnect resistance. Moreover, all the other cells are at LRS, which induce a large current through sneak paths.

In 3D-ICML, the impact of cell locations is also determined by the length of CLs. When its crossbar array has a size of $\mathrm{N} \times \mathrm{N}$, the longest CL goes through N RRAM devices, and hence, the cell locations shows a compatible impact as that in 3D-HIM. However, most of CLs are shorter and have better SM. In other words, compared to 3D-HIM, the SM in the worst-case scenario remains the same, but the one in the best-case scenario improves significantly. Consequently, the SM difference between the worst-case and the best-case cell locations becomes much larger, which is demonstrated by the simulation results of a four-layer 3D-ICML in Fig. 8(d).

C. Read Operation

The grow of crossbar array capacity is primary constrained by read operation performance, *i.e.*, sensing margin and sensing speed. In this subsection, we will analyse the impacts of the data pattern, the cell location, and the array size on SM of the proposed designs.

1) SM under Different Memory Configurations: Fig. 8(b) compares the SMs of the conventional 3D structure (in *blue* curves) and 3D-HIM (in *black* curves) under the different memory configuration. The worst-case cell location and data pattern are applied here.

Since each memory layer in the conventional 3D structure are completely isolated from others, the SM is determined only by one crossbar array and is not affected by the layer numbers. The control signals in 3D-HIM, *e.g.*, WLs and BLs, drive twice number of memory cells as the conventional design, which result in a bigger sneak path current. Therefore, 3D-HIM loses $10 \sim 20\%$ in SM compared to the conventional 3D RRAM. However, further stacking more layers only induces slight SM degradation because of the interleaved device structure.

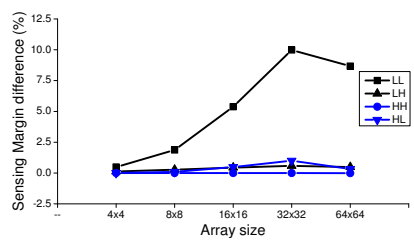
For both the conventional design and 3D-HIM, the SM decreases significantly as increasing the array size. When the size of the crossbar array is 32×32, 3D-HIM obtains an SM of $\sim 20\%$, which is sufficiently large for data detection in read operations.

As we stated in Section V-B2, the worst-case scenario in 3D-ICML are 3D-HIM are quite similar. Hence, the simulation result of 3D-ICML in Fig. 8(e) looks alike to Fig. 8(b).

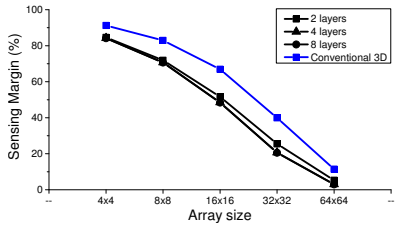
Fig. 9 presents the SM of 3D-HIM with respect to the capacity of a memory island. For instance, a 2048-bit memory can be constructed in the form of $32\times32\times2$ or $16\times16\times8$. The SM of the 8-layer design is 23% more than the SM of the 2-layer design. Furthermore, more memory layers also improves data throughput: the read bandwidth of the 2-layer or 8-layer design is 32 or 64 bits, respectively. In summary, to improve memory capacity, stacking more layers is more beneficial than increasing the crossbar array size in each layer.

2) Impact of Sneak Path on Sensing Voltage: Again, we use 3D-HIM with four layers as the example to thoroughly analyse the SM degradation while increasing crossbar array size. Fig. 10 shows the sensing voltage across $\mathrm{R_{sense}}$. In the ideal condition without considering IR, the sensing current degrades no matter the target memory cell is at HRS or LRS, as the *blue* and red curves showed in Fig. 10, respectively. Enlarging the array size results in more leakage from the leakage paths, which is an innate dilemma of the crossbar array.

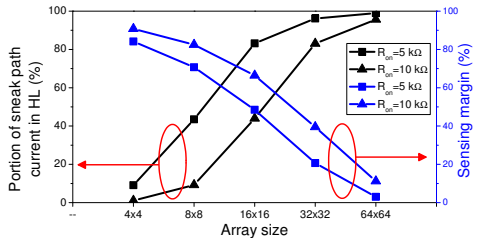
The *black* curves in Fig. 10 are simulation results under four data patterns in the non-ideal condition (that is, including the impact of IR). The SM difference under the ideal and non-ideal conditions is caused by the sneak path conducting current. As crossbar array size increases, the growth of the sneak path conducting current pushes the sensing voltage across $\mathrm{R_{sense}}$ away from its ideal value. The simulation results also show that the enlarge of array size, equal to increase IR due to longer length of driving path, makes the sense voltages in LH, HL and HH patterns increase and the sense voltage in



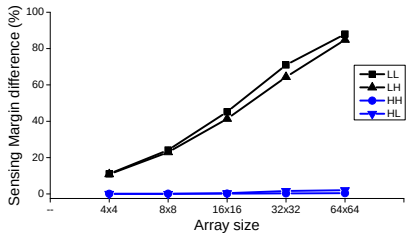
(a) 3D-HIM: SM difference between the worst-case and the best-case cell locations in a four-layer structure, under four typical data patterns.



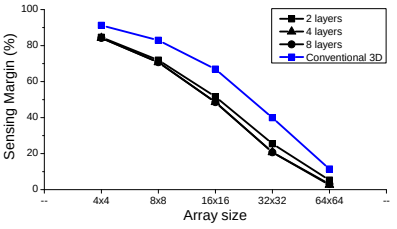
(b) 3D-HIM: SM in the worst-case condition.



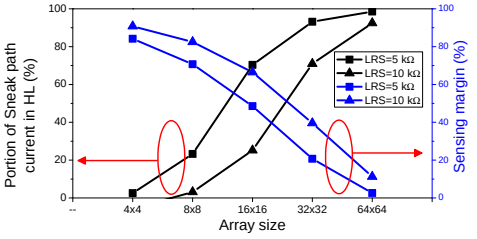
(c) 3D-HIM: The ratio of the sneak path conducting current over the sensing current through R_{sense} and SM of a four-layer structure under various R_{on}.



(d) 3D-ICML: SM difference between the worst-case and the best-case cell locations in a four-layer structure, under four typical data patterns.



(e) 3D-ICML: SM in the worst-case condition.



(f) 3D-ICML: The ratio of the sneak path conducting current over the sensing current through R_{sense} and SM of a four-layer structure under various R_{on}.

Fig. 8. Simulation results of 3D-HIM and 3D-ICML.

LL pattern reduce. Since we require a high sensing voltage to detect the LRS in LH or LL pattern, and a low sensing voltage to detect the HRS in HL or HH pattern, the sneak path conducting currents have a negative impact on LL, HL, or HH data patterns and a positive impact on LH data pattern.

The worst-case scenario happens in LL and HL data patterns. As the crossbar array size increases and the IR grows, the voltage difference between LL and HL reduces significantly. In the LL data pattern, the IR increase causes a relatively small difference compared to ideal curve and results in less SM degradation. However, the sensing voltage in the HL data pattern can increase ten times of that in ideal condition, which significantly degrades SM. Compared to HL, other data patterns have relatively less impact on sensing voltage in non-ideal condition.

Fig. 11 shows the sneak path conducting voltage of the same 3D-HIM design, which is defined as the difference of the sensing voltage across R_{sense} in the ideal and the non-ideal conditions. Let's observe the worst-case scenario – LL and HL data patterns, only. The sneak path increases the current under the HL data pattern, but decrease the current under the LH data pattern. Thus, the sensing margin reduces. Moreover, the simulation result shows that the sensing current under the LL data pattern in 64×64 is already dominated by the sneak path conducting current. This explains the inconsistency of the SM difference in LL pattern in Fig. 8(a): the effect of the sneak conducting current becomes stronger than the location difference with the array expansion.

Fig. 8(c) shows the composition of the sensing current under 'HL' pattern of 3D-HIM structure. As the array size increases, the percentage of the sneak path conducting current raises. In a $64\times64\times4$ 3D-HIM, the conducting current in the sneak path contributes 99% of the sensing current in 'HL' data pattern, which makes it impossible to detect the correct memory status.

Increasing R_{on} can dramatically suppress sneak path current and relieve its impact. For example, increasing R_{on} from $5K\Omega$ to $10K\Omega$ can eliminate 32% and 4.7% of the sneak path conducting current in a 16×16 and 64×64 array, respectively. Correspondingly, their sense margins improve 20.5% and 11%, respectively.

Fig. 8(f) shows the simulation results of 3D-ICML structure under the same configuration. In 3D-ICML, the sneak path conducting current under 'HL' pattern is has less effect than that of the 3D-HIM. Substituting the RRAM cells with high R_{on} devices can also effectively improve the performance of read operations.

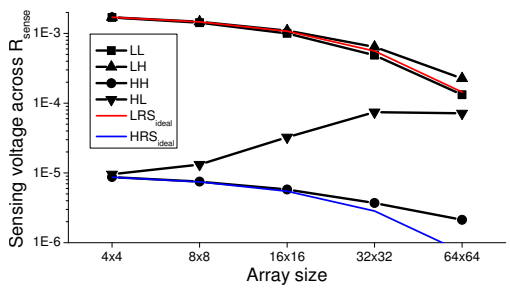


Fig. 10. Then sensing voltage across R_{sense} in 3D-HIM.

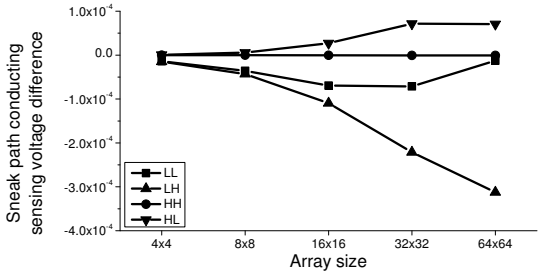


Fig. 11. The sneak path conducting voltage difference in 3D-HIM.

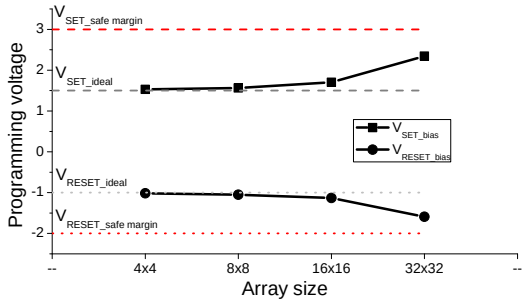


Fig. 12. Proper programming voltage

D. Write operation

The cell location and the data pattern also affect the write operations. The worst case happens at the same location and under the same data pattern in the read operation. Due to space limitation, we only discuss the worst case scenario and follow the explanation for read operation.

Enlarging array size of 3D-HIM increases the total IR in a driving path. To compensate the impact of the increasing voltage drop on the IR and properly program the target cells, a higher bias between WLs and BLs (V_{SET} or V_{RESET}) is required. The corresponding simulation for a four-layer 3D-HIM with various array sizes is shown in Fig. 12. The result of 3D-ICML is similar and omitted here due to space limitation.

The two dotted *gray* lines represent the required SET and RESET voltages across a RRAM cell, which are exactly $V_{SET-bias}$ and $V_{RESET-bias}$ in the ideal condition. However, the impact of the IR cannot be ignored in a real design and it results in the increase of programming voltages as array size increases as demonstrated with the *black* curves. The dotted *red* lines constrain the safe margins of programming voltages, which double the range of the *gray* curves. If V_{SET} or V_{RESET} exceeds the safe margins, some unselected cells may be overwritten since their voltage drops are higher than the threshold. As a result, the proper programming voltages (*black* curves) and safe programming margins (*rea* lines) confine the array size. Our simulation shows that the maximal allowable array size of 3D-HIM is 32×32 to satisfy the constraints in write operations.

E. Discussions

The simulation results show that the two proposed structures, 3D-ICML and 3D-HIM, can obtain the similar performance in the worst-case condition. However, 3D-ICML is more advantageous in the other operation conditions when the CL driving path has a smaller length. The reference cell design presented by Chen *et. al.* [29] can also be utilized along the longest driving path to alleviate the sensing margin degradation. Such a design might benefit 3D-ICML more than 3D-HIM.

Alignment among multiple layers increases process complexity, which could potentially be an issue in 3D structures requiring direct signal transportation from the bottom layer to the top layer. However, our proposed design does not have such direct signal transportation, and therefore, only the relative alignment between the adjacent layers is important. In other words, alignment mismatch does not accumulate along the stacking layers [19]. Hence, our designs alleviate the process difficulty in 3D stacking structure.

Our simulations also show that the RRAM device with higher LRS advances in power consumption and sensing margin. Such high LRS device can be achieved by shrinking device footprint or increasing thickness of oxide materials, which has been thoroughly studied in previous researches [30].

VI. Conclusion

In this paper, we propose two 3D stacking structures built upon bipolar RRAM crossbar arrays, called as 3D-ICML and 3D-HIM. The proposed designs are formed by alternating the deposition of RRAM materials in forward and reverse sequences. As demonstrated by the simulation results, the interleaved device structure help maintain sensing margin and proper programming voltage while suppressing impact of sneak paths and leakage current. Compared to other conventional 3D RRAM structures, the proposed designs have advantages in memory capacity without losing performance. Intuitively, both 3D-ICML and 3D-HIM can be utilized in any bipolar RRAM, especially those materials with a higher R_{on} are preferable. The proposed structures are expected to be more promising in high density non-volatile memory system for future mass storage.

Acknowledgements

Contractor acknowledges Government support in the publication of this paper. This material is based upon work funded by AFRL under contract No. FA8750-11-2-0046. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of AFRL.

References

[1] ITRS, "International Technology Roadmap for Semiconductors 2011 Edition," 2011, http://www.itrs.net/Links/2011ITRS/Home2011.htm.

[2] C. Chevallier, C. Siau, S. Lim, S. Namala, M. Matsuoka, B. Bateman, and D. Rinerson, "A 0.13μm 64Mb multi-layered conductive metal-oxide memory," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*. IEEE, 2010, pp. 260–261.

[3] H. Chung, B. Jeong, B. Min, Y. Choi, B. Cho, J. Shin, J. Kim, J. Sunwoo, J. Park, Q. Wang *et al.*, "A 58nm 1.8 v 1gb pram with 6.4 mb/s program bw," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2011, pp. 500–502.

[4] D. Halupka, S. Huda, W. Song, A. Sheikholeslami, K. Tsunoda, C. Yoshida, and M. Aoki, "Negative-resistance read and write schemes for STT-MRAM in 0.13μm CMOS," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2010, pp. 256–257.

[5] I. Baek, D. Kim, M. Lee, H. Kim, E. Yim, M. Lee, J. Lee, S. Ahn, S. Seo, J. Lee *et al.*, "Multi-layer cross-point binary oxide resistive memory (OxRRAM) for post-NAND storage application," in *IEEE International Electron Devices Meeting (IEDM)*, 2005, pp. 750–753.

[6] R. Waser and M. Aono, "Nanoionics-based resistive switching memories," *Nature Materials*, vol. 6, no. 11, pp. 833–840, 2007.

[7] L. Chua, "Resistance switching memories are memristors," *Applied Physics A: Materials Science & Processing*, pp. 1–19, 2011.

[8] J.P. Strachan, et al., "The switching location of a bipolar memristor: chemical, thermal and structural mapping," *Nanotechnology*, vol. 22, p. 254015, 2011.

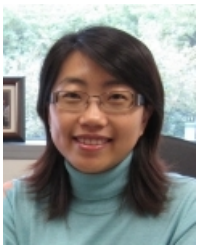
[9] L. Chua, "Memristor-the missing circuit element," *IEEE Transactions on Circuit Theory*, vol. 18, no. 5, pp. 507–519, 1971.

[10] D. Lewis and H. Lee, "Architectural evaluation of 3D stacked RRAM caches," in *IEEE International Conference on 3D System Integration (3DIC)*, 2009, pp. 1–4.

[11] M. Meier, R. Rosezin, S. Gilles, A. Rudiger, C. Kugeler, and R. Waser, "A multilayer RRAM nanoarchitecture with resistively switching Ag-doped spin-on glass," in *IEEE 10th International Conference on Ultimate Integration of Silicon (ULIS)*, 2009, pp. 143–146.
[12] A. Madayag and Z. Zhou, "Optimization of spin-on-glass process for multilevel metal interconnects," in *University/Government/Industry Microelectronics Symposium, 2001. Proceedings of the Fourteenth Biennial*. IEEE, 2001, pp. 136–139.
[13] R. Soni, P. Meuffels, H. Kohlstedt, C. Kugeler, and R. Waser, "Reliability analysis of the low resistance state stability of $Ge_{0.3}Se_{0.7}$ based solid electrolyte nonvolatile memory cells," *Applied Physics Letters*, vol. 94, no. 12, p. 3503, 2009.
[14] E. Linn, R. Rosezin, C. Kugeler, and R. Waser, "Complementary resistive switches for passive nanocrossbar memories," *Nature Materials*, 2010.
[15] H. Li and Y. Chen, "An overview of non-volatile memory technology and the implication for tools and architectures," in *IEEE Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2009, pp. 731–736.
[16] Y. Hosoi, Y. Tamai, T. Ohnishi, K. Ishihara, T. Shibuya, Y. Inoue, S. Yamazaki, T. Nakano, S. Ohnishi, N. Awaya *et al.*, "High speed unipolar switching resistance RAM (RRAM) technology," in *IEEE International Electron Devices Meeting (IEDM)*, 2006, pp. 1–4.
[17] R. Soni, M. Meier, A. Rudiger, B. Hollander, C. Kugeler, and R. Waser, "Integration of Ge_xSe_{1-x} in crossbar arrays for non-volatile memory applications," *Microelectronic Engineering*, vol. 86, no. 4-6, pp. 1054–1056, 2009.
[18] S. Jo, K. Kim, and W. Lu, "High-density crossbar arrays based on a Si memristive system," *Nano letters*, vol. 9, no. 2, pp. 870–874, 2009.
[19] D. Strukov and R. Williams, "Four-dimensional address topology for circuits with stacked multilayer crossbar arrays," *Proceedings of the National Academy of Sciences*, vol. 106, no. 48, p. 20155, 2009.
[20] C. Kugeler, M. Meier, R. Rosezin, S. Gilles, and R. Waser, "High density 3D memory architecture based on the resistive switching effect," *Solid-State Electronics*, vol. 53, no. 12, pp. 1287–1292, 2009.
[21] Y. Lu, "3D technology based circuit and architecture design," in *IEEE International Conference on Communications, Circuits and Systems*, 2009, pp. 1124–1128.
[22] T. Gao, B. Coenegrachts, J. Waeterloos, G. Beyer, H. Meynen, M. Van Hove, and K. Maex, "Integration of unlanded via in a non-etchback SOG direct-on-metal approach in 0.25 micron CMOS process," in *IEEE Proceedings of the IEEE International Interconnect Technology Conference*, 1998, pp. 45–47.
[23] K. Saraswat, K. Banerjee, A. Joshi, P. Kalavade, P. Kapur, and S. Souri, "3-D ICs: Motivation, performance analysis, and technology," in *IEEE Proceedings of the 26th European Solid-State Circuits Conference (ESSCIRC)*, 2000, pp. 406–414.
[24] M. Johnson, A. Al-Shamma, D. Bosch, M. Crowley, M. Farmwald, L. Fasoli, A. Ilkbahar, B. Kleveland, T. Lee, T. Liu *et al.*, "512-Mb PROM with a three-dimensional array of diode/antifuse memory cells," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 11, pp. 1920–1928, 2003.
[25] C. Nauenheim, C. Kugeler, A. Rudiger, R. Waser, A. Flocke, and T. Noll, "Nano-crossbar arrays for nonvolatile resistive RAM (RRAM) applications," in *IEEE 8th IEEE Conference on Nanotechnology*, 2008, pp. 464–467.
[26] Y. Li, S. Long, M. Zhang, Q. Liu, L. Shao, S. Zhang, Y. Wang, Q. Zuo, S. Liu, and M. Liu, "Resistive Switching Properties of $Au/ZrO_2/Ag$ Structure for Low-Voltage Nonvolatile Memory Applications," *IEEE Electron Device Letters*, vol. 31, no. 2, pp. 117–119, 2010.
[27] C. Cheng, A. Chin, and F. Yeh, "Novel Ultra-low power RRAM with good endurance and retention," in *IEEE 2010 Symposium on VLSI Technology (VLSIT)*, 2010, pp. 85–86.
[28] J. Liang and H. Wong, "Cross-Point Memory Array Without Cell Selectors – Device Characteristics and Data Storage Pattern Dependencies," *IEEE Transactions on Electron Devices*, vol. 57, no. 10, pp. 2531–2538, 2010.
[29] Y.-C. Chen, W. Zhang, and H. Li, "A Look Up Table Design with 3D Bipolar RRAMs," in *Asia and South Pacific Design Automation Conference (ASPDAC)*, 2012, pp. 73–78.
[30] B.J. Choi, et al., "Purely Electronic Switching with High Uniformity, Resistance Tunability, and Good Retention in Pt-Dispersed SiO2 Thin Films for ReRAM," *Advanced Materials*, 2011.

Yi-Chung Chen received the B.S. degree in nanotechnology from Yuan Ze University, Taoyuan, Taiwan, in 2006.

He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University. His research interests include architecture/circuit/device co-design for emerging nonvolatile memory system, FPGA, embedded system and reconfigurable computing.



Hai (Helen) Li (M08) received the B.S. and M.S. degrees in microelectronics from Tsinghua University, Beijing, China, and the Ph.D. degree from the Electrical and Computer Engineering Department at Purdue University, West Lafayette, IN, in 2004.

She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, Brooklyn. She was with Qualcomm, Intel, and Seagate earlier. Her research interests include architecture/circuit/device co-optimization for green VLSI systems, emerging embedded memory design, 3-D integration technology and design, design for new devices, and neuromorphic architecture for brain-inspired electronic systems. She has published more than 70 technical papers in refereed journals and conferences, filed 59 U.S. patents (41 granted), and one Seagate Trade Secret. She also authored one book chapter and will publish one book about emerging memory design by CRC Press.

Dr. Li received two best paper awards and three best paper nominations from ISQED, ISLPED, DATE, and ASPDAC. She has served as a technical program committee member of more than 20 international conference series. She has been a reviewer for numerous journals and conferences.

Wei Zhang (M05) received the B.S. and M.S. degrees in electrical engineering from Harbin Institute of Technology, Harbin, China, in 1999 and 2001, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 2009.

She is currently an Assistant Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. Her current research areas include multicore systems, reconfigurable computing, and emerging technology.

Dr. Zhang has published more than 30 technical papers in referred journals and conferences, and authored three book chapters. She won a research Prize in Princeton and a best paper award. Dr. Zhang has served in many conferences and journals as technical program committee and reviewer.

Robinson E. Pino (S99-M09-SM10) received the B.E. degree in Electrical Engineering with honors, summa cum laude, from the City University of New York, City College, NY in 2002. Dr. Pino received the M.Sc. and Ph.D. degrees in Electrical Engineering from Rensselaer Polytechnic Institute, Troy, NY in 2003 and 2005 respectively.

Since 2012, Dr. Pino is a Senior Scientist at ICF International advancing the state of the art in cyber security for the United States Department of Defense. From 2009 to 2012, Dr. Pino was Senior Electronics Engineer at the United States Air Force Research Laboratory where he was Program Manager and Principle Scientist for the Computational Intelligence and Neuromorphic research efforts.

In Industry, Dr. Pino worked at IBM, 2005 to 2009, as Advisory Scientist/Engineer enabling advanced CMOS technologies, and as Business Analyst within IBM's Photomask business unit, and from 2006 to 2009 served as adjunct professor at the University of Vermont. Dr. Pino is recipient of numerous awards and professional distinctions, has published over 40 technical papers, two books, and holds two patents, five pending.